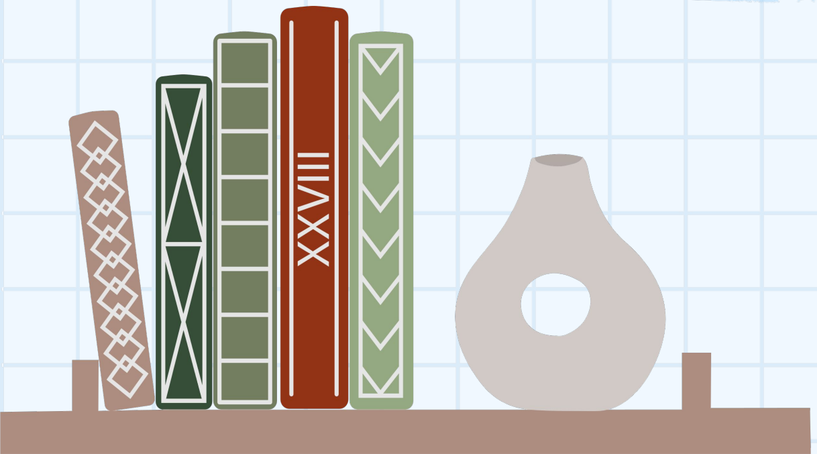
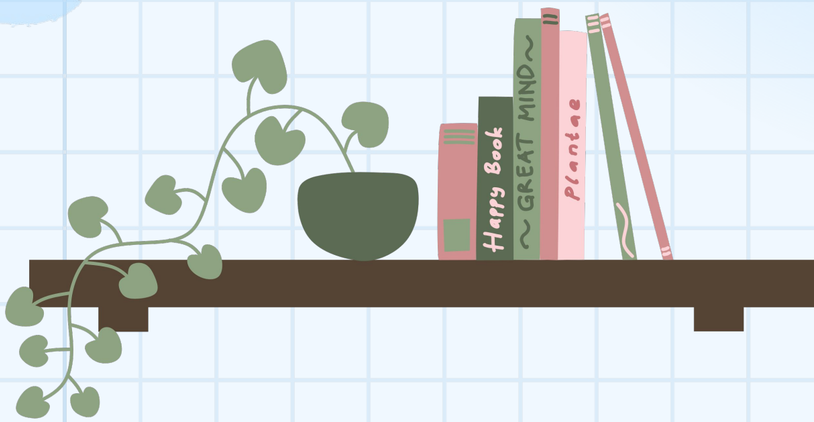
Powered by
Futurense

BS./BSC.IN

Applied ai and Data Science

Basics of Data Analytics





Powered by Futurense

Let's dive into and learn:

1. Combining datasets

Combining Data

Powered by

Futurense

- While working with data, in the data wrangling stage you will often have to combine different datasets.
- Rarely will your data project involve working with a single dataset
- Relational datasets can be put together through various ways

Keys

- Need to understand keys that are used to combine datasets together
- Every join involves a pair of keys
- In relational databases, primary keys can be used as fields to join tables on

Keys

Powered by

Futurense

Primary Key

- a variable that uniquely identifies each observation

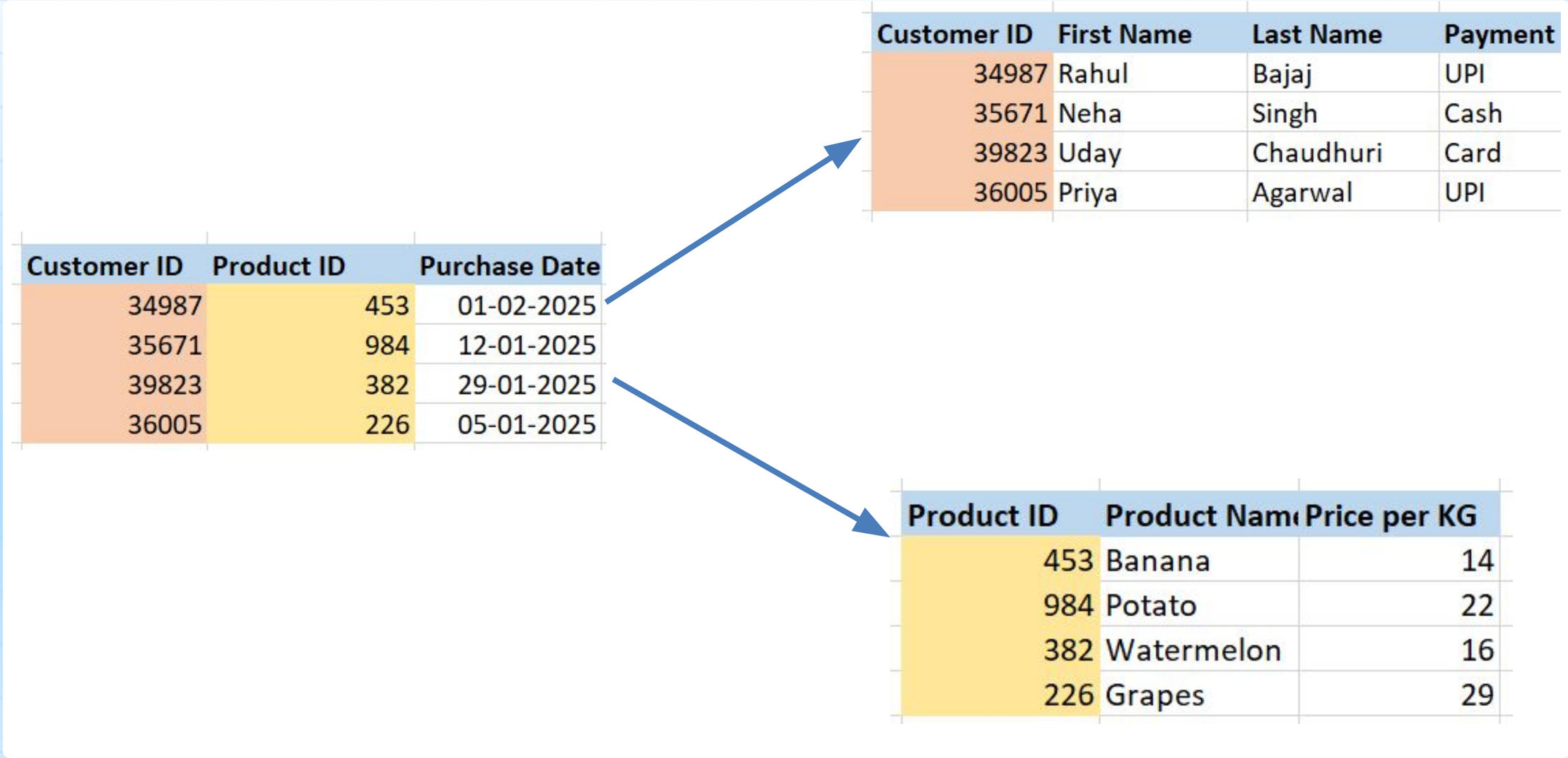
Compound Key

- when more than one variable is needed to uniquely identify an observation

Foreign Key

- a field in a table which references the primary key of another table.
- In a relational database, one way to join two tables is by connecting

Keys



Customer ID	First Name	Last Name	Payment
34987	Rahul	Bajaj	UPI
35671	Neha	Singh	Cash
39823	Uday	Chaudhuri	Card
36005	Priya	Agarwal	UPI

Customer ID	Product ID	Purchase Date
34987	453	01-02-2025
35671	984	12-01-2025
39823	382	29-01-2025
36005	226	05-01-2025

Product ID	Product Nam	Price per KG
453	Banana	14
984	Potato	22
382	Watermelon	16
226	Grapes	29

Types of Relationships

Powered by Futurense

One-to-one relationships

- When a one-to-one relationship exists between two tables, a given record in one table is uniquely related to exactly one record in another table.

One-to-many relationships

- In a one-to-many relationship, a record in one table can be related to one or more records in a second table. However, a given record in the second table will only be related to one record in the first table.

Types of Relationships

Powered by

Futurense

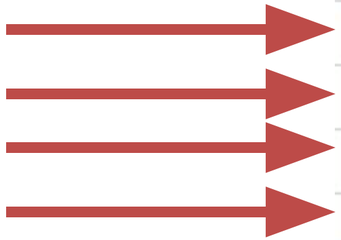
Many-to-many relationship:

- In a many-to-many relationship, records in a given table A can be related to one or more records in another table B , and records in table B can also be related to many records in table A.

One –to –One Relationships

Powered by Futurense

Customer ID	Purchase Date	Purchase Date
34987	01-02-2025	01-02-2025
35671	12-01-2025	12-01-2025
39823	29-01-2025	29-01-2025
36005	05-01-2025	05-01-2025



Customer ID	First Name	Last Name	Payment Mode
34987	Rahul	Bajaj	UPI
35671	Neha	Singh	Cash
39823	Uday	Chaudhuri	Card
36005	Priya	Agarwal	UPI

One –to – Many Relationships

Powered by Futurense

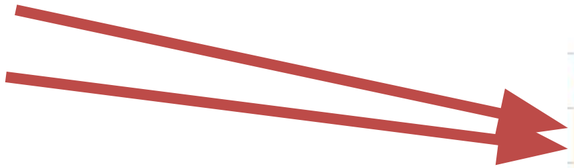
Transaction ID	Purchase Date
53666	01-02-2025
54333	12-01-2025
56473	29-01-2025
56234	05-01-2025

Transaction ID	Product ID	Product Name	Price per KG
53666	453	Banana	14
53666	984	Potato	22
53666	382	Watermelon	16
53666	226	Grapes	29
54333	984	Potato	22
54333	382	Watermelon	16
56473	382	Watermelon	16
56473	226	Grapes	29
56234	453	Banana	14
56234	984	Potato	22
56234	382	Watermelon	16

Many –to-One Relationships

Powered by Futurense

Purchase Date	Transaction ID	Customer ID
01-02-2025	57334	34987
02-02-2025	53886	34987
03-02-2025	55227	35671
04-02-2025	51348	35671
05-02-2025	50872	35671
06-02-2025	53666	39823
07-02-2025	54333	39823
08-02-2025	56473	36005
09-02-2025	56234	36005



Customer ID	First Name	Last Name	Payment Mode
34987	Rahul	Bajaj	UPI
35671	Neha	Singh	Cash
39823	Uday	Chaudhuri	Card
36005	Priya	Agarwal	UPI

Recap

Powered by Futurense

Combining Datasets together

Powered by Futurense

Thank you